

```

1 TTCGCCTGCG GGCCGGCACT GCTCACCTCT CGTCCAGGGA CATGACGGGC
51 ACGCCAGGCG CCGTTGCCAC CCGGATGGC GAGGCCCCCG AGCGCTCCCC
101 GCCCTGCACT CCGAGCTACG ACCTCACGGG CAAGGTGATG CTTCTGGGAG
151 ACACAGGCGT CGGCAAAACA TGTTTCCTGA TCCAATTCAA AGACGGGGCC
201 TTCCTGTCCG GAACCTTCAT AGCCACCGTC GGCATAGACT TCAGGAACAA
251 GGTGGTGACT GTGGATGGCG TGAGAGTGAA GCTGCAGATC TGGGACACCG
301 CTGGGCAGGA ACGGTTCCGA AGCGTCACCC ATGCTTATTA CAGAGATGCT
351 CAGGCTTGC TTCTGCTGTA TGACATCACC AACAAATCTT CTTTCGACAA
401 CATCAGGGCC TGGCTCACTG AGATTCATGA GTATGCCCAG AGGGACGTGG
451 TGATCATGCT GCTAGGCAAC AAGGCGGATA TGAGCAGCGA AAGAGTGATC
501 CGTTCCGAAG ACGGAGAGAC CTTGGCCAGG GAGTACGGTG TTCCCTTCTT
551 GGAGACCAAG GCCAAGACTG GCATGAATGT GGAGTTAGCC TTTCTGGCCA
601 TCGCCAAGGA ACGCAATAAC CCGGCCGGGC ATCAGGCGGA TGAGCCACG
651 TTCCAGATCC GAGACTATGT AGAGTCCCAG AAGAAGCGCT CCAGCTGCTG
701 CTCCTTCATG TGAATCCCAG GGGGCAGAGA GGAGGCTCTG GAGGCACACA
751 GGATGCAGCC TTCCCCCTCC CAGGCCTGGC TTATTCCAAG AGGCTGAGCC
801 AATGGGGAGA AAGATGGAGG ACTCACTGCA CAGCCGCTTC CTAGCAGGGA
851 GCTATACTCC AACTCCTACT TGAGTTCCTG CCGTCTCCCC GCATCCACAG
901 GGAGGGTAAA ACCTTAGCT TTTATTTTAA TAGTACATAA TTTAATACCA
951 AAAAAGGCGC CTGGATCCCC AAAAAACCGA GGCTGGGAGC TAGTGGCCCT
1001 TTTGCTTTCT AGGACTTGGG GGGCCGGCCC TCCCTCCTAA GCATAACAAA
1051 GGTGGTGTG CTCCAGCTCA GCCCCAGGGG ACACAGATGC ACTTTGGGGG
1101 TGAGGGCAGG TAATGACTCC ATCGCACCTT CAGTTCAGCT GGACAGAGGC
1151 TCAGGTGACC CCAGCCTTCA CTGTCTCCCG CTCTCCAGGA GCTTATCTTC
1201 GCCCCTCTC CCAAATAAGT GGGCCCTTGT GCTGTGAGGA AGACCAAAGC
1251 CTCAGGGAAG ATAAGAGATA TGGAGATGGG AGGGGGAGGA CAAGGGGCAG
1301 AGAGTAGGGT CTAGCTGGCT ATCTCTGGCC TTAGTAACAC CCCCTGGAG
1351 GCATGCCCTT TTTCTCCAGC ACACAAGCAC ATTGGGGCAC CTGGAAATAT
1401 TGGTTCAGG CTCCTGTTCT CTGGAATTCA GATCCTGGGG GAGCCCCCTC
1451 CCCCCCTGAA TCCCTGGCTT AGCTACCTTC CTGCCTGTGC ACCTAAAAAC
1501 CTCAGGTGAG AACTAGGAAA AGAGTTTGT TTTTATTTT TTGAAATGGA
1551 GTCTCGTTCT GTCCGCCAGG CTGAGGTGCA GTAGTGCAAT CTCGCTCAC
1601 TACAACCTCC ACTCCCTGGG GCTCAAGCGA TCCTCCACAC TCAGCCGCGC
1651 AAGTAGCTGG GACTATAGGT GTGTACCATC ACACCTGGCT AATTTTTGTA
1701 TTTTTTTGTA ACACAGGGTT TCGCCATGTT GCCCAGGCTG GTCTTGAATT
1751 CCTGAGCTCA AGCAACCTGC CCGCCTCGGC CTCCCAAAGT ACTGGGATTA
1801 CACGCAGAAG GCACCATGCC CAGGCTAGAT GTGTCTTATC CCAATCCTTT
1851 GGCAGGCATG CAGCTCCACA GCGGATTCTT TCAAGCAGCT GAAGTGTTTA
1901 GCCCTCCTGG GTTAAGAGCC AGATAAGGAG AAATCCCTTT CCTAGGTTTG
1951 GAATGTGTTG TGAAAAAATA GAGAAATCCC TGGCTCCTGG AGCTGGTGGG
2001 AGACAAGATT AAGCAAACCT CCCCTGACAT GTATCCCTTT GACCCCAAGC
2051 TCTGCCTCCT CCCTGACCAC CCATGCCCTT TCCTTTAACT TCTCAAACAG
2101 ATACCAGGGC CTAAACTGCT TTACCTCCCC TCCTACTGAG TCAGGTTAGG
2151 TGGTGGGAGG TCACCCATTT CCGAGTTAAA CCAATGCAAT ATGAGTAAAA
2201 CAAAGTCATG TGGGTATGTC TGGGGTAGAG AGAGGGGTAG CAAGTTCATG
2251 TGTCTCCTT GGTACATAT CTCCCAAAGC TCTGATCCCT GCCATGGGAA
2301 GTGGACAGGA AACATGAGGT CATGACCTGC AGGCATCTTT ACTGCAGCTC
2351 TGCCGGCCTG GAGGGGGAGA GGGGGAGGAA GAAGTATGCG CTGCACATTT
2401 CTGAGGCTAC TGCATTTGCT TTCAAGGCAG AAATCTTGCT CTGAGCAGTC
2451 AGCGGCTCCA GTTTGGGCCC GATAAGGAAG TTCTCCGTGG CCTCCCTCAG
2501 GCAGAGCAGG GAGGAGGCTG ACATTGCCAG TCTCTTCTGG GGCCCAAGGC
2551 AGGTTGCAGG AGATCCAATC CCATAGACAG CTCTGGGCCT CTGCAATTTG
2601 AGTTTTTTCAG AATTAACTG CAGTATTTTG GAAAGCAAAA AAAAAAAAAA
2651 AAAAAAAAAA AAAAAAAAAA AAAA (SEQ ID NO:1)

```

FEATURES:

5'UTR: 1-41
Start Codon: 42
Stop Codon: 711
3'UTR: 714

Homologous proteins:

Top 10 BLAST Hits

	Score	E
CRA 103000001517087 /altid=gi 10946770 /def=ref NP_067386.1 RA...	425	e-117
CRA 1000682330460 /altid=gi 7657492 /def=ref NP_055168.1 RAB26...	297	4e-79
CRA 18000004977238 /altid=gi 1710022 /def=sp P51156 RB26_RAT RA...	294	3e-78
CRA 18000005013109 /altid=gi 1083775 /def=pir JC2528 GTP-bindi...	293	7e-78
CRA 89000000198627 /altid=gi 7296421 /def=gb AAF51708.1 (AE003...	273	9e-72
CRA 18000005076419 /altid=gi 7438397 /def=pir T15123 hypotheti...	207	4e-52
CRA 18000004912300 /altid=gi 134236 /def=sp P20791 SAS2_DICDI G...	203	7e-51
CRA 980000043536338 /altid=gi 12963499 /def=ref NP_075615.1 cel...	203	9e-51
CRA 18000004929618 /altid=gi 131798 /def=sp P24407 RAB8_HUMAN R...	202	1e-50
CRA 18000004952869 /altid=gi 131848 /def=sp P22128 RAB8_DISOM R...	202	2e-50
CRA 18000005221564 /altid=gi 4586580 /def=dbj BAA76422.1 (AB02...	202	2e-50

BLAST dbEST hits:

	Score	E
gi 13033710 /dataset=dbest /taxon=960...	1318	0.0
gi 12785775 /dataset=dbest /taxon=960...	1316	0.0
gi 12904236 /dataset=dbest /taxon=960...	1035	0.0
gi 9093496 /dataset=dbest /taxon=9606...	694	0.0

EXPRESSION INFORMATION FOR MODULATORY USE:

library source:

From BLAST dbEST hits:

gi|13033710 prostate
gi|12785775 brain
gi|12904236 T cells from T cell leukemia
gi|9093496 leukopheresis

From tissue screening panels:

leukocyte

```

1 MTGTPGAVAT RDGEAPERSP PCSPSYDLTG KVMLLGDGTG GKTCFLIQFK
51 DGAFLSGTFI ATVGIDFRNK VVTVDGVRVK LQIWDTAGQE RFRSVTHAYY
101 RDAQALLLLY DITNKSSFDN IRAWLTEIHE YAQRDVVIML LGNKADMSSE
151 RVIRSEGET LAREYGVFFL ETSAKTGMNV ELAFLAIAKE LKYRAGHQAD
201 EPSFQIRDYV ESQKRSSCC SFM (SEQ ID NO:2)

```

FEATURES:

Functional domains and key regions:

[1] PDOC00001 PS00001 ASN_GLYCOSYLATION
N-glycosylation site

114-117 NKSS

[2] PDOC00004 PS00004 CAMP_PHOSPHO_SITE
cAMP- and cGMP-dependent protein kinase phosphorylation site

Number of matches: 2

1 214-217 KKRS
2 215-218 KRSS

[3] PDOC00005 PS00005 PKC_PHOSPHO_SITE
Protein kinase C phosphorylation site

Number of matches: 5

1 29-31 TGK
2 113-115 TNK
3 149-151 SER
4 173-175 SAK
5 212-214 SQK

[4] PDOC00006 PS00006 CK2_PHOSPHO_SITE
Casein kinase II phosphorylation site

116-119 SSFD

[5] PDOC00008 PS00008 MYRISTYL
N-myristoylation site

Number of matches: 5

1 3-8 GTPGAV
2 6-11 GAVATR
3 39-44 GVGKTC
4 52-57 GAFLSG
5 57-62 GTFIAT

[6] PDOC00017 PS00017 ATP_GTP_A
ATP/GTP-binding site motif A (P-loop)

36-43 GDTGVGKT

[7] PDOC00579 PS00675 SIGMA54_INTERACT_1
Sigma-54 interaction domain ATP-binding region A signature

32-45 VMLLGDGTGVGKTCF

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	48	68	0.715	Putative

BLAST Alignment to Top Hit:

>CRA|103000001517087 /altid=gi|10946770 /def=ref|NP_067386.1| RAB37,
member of RAS oncogene family; GTPase Rab37 [Mus
musculus] /org=Mus musculus /taxon=10090 /dataset=nraa
/length=223
Length = 223

Score = 425 bits (1081), Expect = e-117
Identities = 209/223 (93%), Positives = 215/223 (95%)
Frame = +3

Query: 42 MTGTPGAVATRDGEAPERSPPCSPSYDLTGKVMLLGDTGVGKTCFLIQFKDGAFLSGTFI 221
MTGTPGA DGEAPERSPP SP+YDLTGKVMLLGD+GVGKTCFLIQFKDGAFLSGTFI
Sbjct: 1 MTGTPGAATAGDGEAPERSPPFSPNYDLTGKVMLLGD SGVGKTCFLIQFKDGAFLSGTFI 60

Query: 222 ATVGIDFRNKVVTVDGVRVKLQIWDTAGQERFRSVTHAYYRDAQALLLYDITNKSSFDN 401
ATVGIDFRNKVVTVDG RVKLQIWDTAGQERFRSVTHAYYRDAQALLLYDITN+SSFDN
Sbjct: 61 ATVGIDFRNKVVTVDGARVKLQIWDTAGQERFRSVTHAYYRDAQALLLYDITNQSSFDN 120

Query: 402 IRAWLTEIHEYAQRDVVIMLLGNKADMSSERVIRSEDGETLAREYGVFPFLETSAKTGMNV 581
IRAWLTEIHEYAQRDVVIMLLGNKAD+SSERVIRSEDGETLAREYGVFP+ETSAKTGMNV
Sbjct: 121 IRAWLTEIHEYAQRDVVIMLLGNKADVSSERVIRSEDGETLAREYGVPFMETSAKTGMNV 180

Query: 582 ELAFLAIAKELKYRAGHQADEPSFQIRDYVESQKKRSSCCSF 710
ELAFLAIAKELKYRAG Q DEPSFQIRDYVESQKKRSSCCSF+
Sbjct: 181 ELAFLAIAKELKYRAGRQDPDEPSFQIRDYVESQKKRSSCCSFV 223 (SEQ ID NO:4)

Hmmer search results (Pfam):

Model	Description	Score	E-value	N
PF00071	Ras family	306.9	8.4e-90	1
CE00060	CE00060 rab_ras_like	213.3	3.7e-60	1
PF01142	Uncharacterized protein family UPF0024	2.6	3.4	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
CE00060	1/1	31	191 ..	25	193 ..	213.3	3.7e-60
PF01142	1/1	185	201 ..	444	462 .]	2.6	3.4
PF00071	1/1	31	223 .]	1	198 []	306.9	8.4e-90

1 AGGGGAGAGA AAAGACCGCA TACCAGGCCA GGTGCGGTGG CTCACGCTTG
51 TAATCCCAGC AATTGGAAG GCCAAGGCAG GCGTATCGCC TGAGGTCAGC
101 AGTTCCAAAC CAGCCTGTCC AACATGGTGA AGTTCTCTAC TAAGAATACA
151 AAAATTACCC AGGCGTGGTG GCGTGACCT GTAGTCCCAG CTGCTCCAGA
201 GGCTGAGGCA GGAGAATTGC TTGAACCTGG GAGGCAGAGG CTGCAATGCG
251 CCAAGATCCC GCCACTGCAC TCCAGCCTGG GCGACAGAGT GAGACTCCGT
301 CTCCGGGAGC CCACGGCATT GAGCAAACCT CGGCATTATT TGCAGCAAGA
351 GCCTCTGGCA TCCAAATAGC AACCAACACC ACGCCTCTGT AGTGTGCTGC
401 GCAGCTCCA CACTCCAGTC TGAGGCTCCC TGTTTGAGTC CCGCCCTATG
451 CCCAGCTGAG GTTATAGCAC GCTCACCTCC AGAAGAGGTA ACCCAAGCTC
501 TTTACTCTAC TGGAGATCAC CTCTGTCCCC ACTCTGGGCG CTTCTCCAG
551 CTGACAGAAA ATACCTCCAG CTGATGTCAG AAAATACAGG GCTGGAGGCT
601 GCGGTAGCAA GTCAATCCCC ACAGGCCTAT GGTGGCCCAT AAGCCACGTC
651 TACCCCTGCT CCTCACCTCC ACACCTAAGT TAAGAATTGC AGGCCGGGCG
701 CAGTGGCTCA CGCCTGTAAT CCCAGCACTT TGGGAGGCTG AGGTGGGCGG
751 ACCGCTGAG GTCAGGAATT TGAGACCAGC TTGGCCAACA TGGCAAAACC
801 CCGTCTCTAC TAAAAATACA AAAAGAAAAA ATAGCCGGGC CTGATGTCGC
851 GCACCTGTAA TCCCAGCTAC TCCGGGAGAC TGAGGCGGGA GTATAGCTTG
901 AACCCGGGAA GCAAAGGTTG CAGTGAGGCG AGATCGCACC ACTGCACTCC
951 AGGCTGGGCG ACAGAGTGAG ACTCTGTCTG AAAAAAAAAA AAAGTGCAGG
1001 TACCCCTCTC CAGCTCTCCC CTCCCTACAC ATCCCTCAAA CCGTCCCGCT
1051 GTAATGCACC CGCCTGTTC CTTGGTAACT TGAAGCTGCT TATAGAATGT
1101 GGAGATGGGG GTAATTGAAA GGTGCGGCCA GGCCACAGAG CCCCTGAGCT
1151 CTGCTACCGG CAACCCAGC TGCACTCCCC ACTCTCTGTC ACCAGGAGCT
1201 GCCGGGTGCC TGGGATATCC TGGCAGCTCT GCTCAAAATG ATCTACGACT
1251 TCATGAATTT ATTTGGCTCC TCCTCGGGGC CAGGGTGAGT GTCATGGGTT
1301 AATAAGGCCG GCCCCGCCCT CAGGAGCGGT CCACTGGGAG ATGTGTGCTG
1351 CGCAGCCCTC TTGCGAAAGC TCTCCCCTGG TGGGACATTC TGGGCACAAC
1401 CAACAGGCCG GGGGAAATGA GAGGTGATCC ATACTAAAGG GTCAAAGTCC
1451 CCGCACCAGG CAGAGGCCCC AAAACACCGC AGCGTACATG TGCTGCAAGG
1501 CGAGTACGGG TTGGTAAACA AAATATATT CAGATGAGCT CGGGCCGGGT
1551 GACTTAACAG ATGAGGAAGT GTCTCGGGGC CATCGGCGGA GGCGCAGCCC
1601 AGGGGTCCCC AGCTCCCCGC CTCGCCACCT GGGGACAGCC CACGGCCCCG
1651 GGCTCGGGCG CCGCCTGCTG TCGCGGTGCG CAGCGACTAC GGGAACCTT
1701 CCGCAGCAGA CGGGGTCCCC GCGGCCCGCT CCCCCAGGGG CAAGCAAGCG
1751 ACCACAGGGG ACCGGTCCCG GGGCTGGATG TGGCTCATGT CCGAAGCGCA
1801 CGGAGCCGAG CCGGTGTTGC TCAGGGAGGC TGCCCGCCCC TTCACGCAGA
1851 CCCTGCGGCT CTGCGTGCCC TCAGGGAACA GCAAGGTCCG AGCCGGTGTG
1901 GTCGAGGGGG GCGAGGAGG GAGGGAGGAG CCTGAGGGGT CCCGGTCCAG
1951 GGAGGGGAGG AGTGGGCGGG GCGGGGTGG GGGCCGTTC CGCGCTCTCC
2001 TTCGCTGCG GGCCGGCACT GCTCACCTCT CGTCCAGGGA CATGACGGGC
2051 ACGCCAGGCG CCGTTGCCAC CCGGGATGGC GAGGCCCCCG AGCGCTCCCC
2101 GCCCTGCACT CCGAGCTACG ACCTCACGGG CAAGGTGGGT GGGCCTCTTC
2151 CGTGAGACCC CCGCCCTCCT CGGCGCTAGC CCCTTCTGG CTGCGTCTGG
2201 GTTGGACTCA GCCCTTCCCC CAGGCAGCTG CGTCTCCAG AGGAGGGAGG
2251 GAGAGAGGGT CAGGACACAG CCTCTGGGC CGTCCAAGC TCTAGGTGTC
2301 TCTGCTGGCT TGGTGGGGGC GGGTCGCGGA AGATCGCAA AACTGAGTGA
2351 TCCCCCGGCC GGCCCCAACT CAGTTCTCTT CTGCCACACT CTGGCAAATA
2401 TGAGCCCCCG GGAGCCCATG CTTCTTGGTG AGGTTAAGC GCGCAACTCT
2451 CGGGGCTCAG GCTGGGAAGG GCTGGGAGAT GGGGACCGAA CGGAGACTCG
2501 GAGAGGACGT CCCCTGCTGG CAGAGGAACT GGCGTTAATG CCATTTTCCG
2551 AGCTAAGCTC TTAGTTGAGA TCTGACATCC AGGTTAAGG CCTGATGTCC
2601 CCCAGCTGCT CCCCTCCCAT TCCACCGCT GGAGGCACTG CCTCCACCT
2651 TCCTCCCTGC AGTCGGAAGC CGCTCCTCCC AGAAGGATGT TGCCAGCCGG
2701 CCTGCAGGTC ACTTGGAAT TTTTGAACC TGAGAAAGAT TTCAGTGGTT
2751 GGTCTTTGCG ATCCCCACT TGAGAGAGCT CCAGGGCTGC TCTCTGGGGC
2801 TTGCTCCCTC TACAGGGGTG TCCTGTATGG AAACAGGTAG GGACAGCAGT
2851 GGAAGTGTCT GTCGCTTCC ATCTGTGTCC TTGGAGTGAG CGGGTACCAG
2901 AAAGTGAAG AACTGCTGAG GGAGCCTAGA GCTTCCACT TTCCTTGCA
2951 GGGTTGGGGA TGGAGTGAG GCTGTCTGG ATTCCGCTGC ATGGCCTGA
3001 AGGAGACCTG CCTCTCTCTG GGCCTCGGTT TCCTCCCCGA CACCAGGGCT
3051 CACCCTTGCT GGGAGCTCA GCCTCCACCC CAGTGTTCG GGGGAAGCCA
3101 CCCTGCAAGT CATCCGCCCA GAGCCGTTGA GATAGGCGTC CTGTGTGGGC

FIGURE 3, page 1 of 7

3151 TTGTGGCAGG AAATGGGCCC CTGCACCCTC GGAGAGGAGG AGCTGCTGTT
3201 GGCCAGGCCC CAGGCTGAGG GGGACTGCCT GACCTTGTG CCCTGCAAAC
3251 CAGCTGGGTT GTTTGCCTAG GAGGTGGCCA GGCTAGGCAG CTGTTTGTGT
3301 TTGGTGGAAT CACCGAGCTG GGTGGGTAGC TGGCATCGTT TGCTCAAGGC
3351 AGCTGTGATC TGTAAGTAC ACAAAGACTG GCCCTCCCTC CCTCCTTCCT
3401 GCTCCAGAGG GTTAAGGAAT TGGAGCCAG GAGTCAGG GCATTCGGCCG
3451 CTCCTATCCC CCACCCCTTC ATCTGTTCCC TGGCCAAGCG GCATTCGGCCG
3501 GAGAGTTGGT CCCAGCCTC CCCGGGCCCTG CCCAGGGGA GTGAGTCCAG
3551 GACCTCTGA GAAAGCCTGG CAGGAGCTCC TTGGACCAGA CTAGGGGTGA
3601 TGTGGCCAC AGGCAGACAG TTCCACCCT GGGCCACTCT TCCCTGGGTC
3651 TTAGGTGATT CACCACGATG ATGGGCCCTA GCCATTAACA GACTCTAGAA
3701 ATACCTCAAA GACATTATCC CTCCTCCTTC TACCCACTAT GGAAACCATG
3751 CCACAGAAAG GTTAAGGAAT CTTCCTAAAG TCACACAGTA GGCCATTTAC
3801 AAATCAAGAC CCATCCTTCA TACCCCTTCT GCTCAGCCAC CCCTGCCTCT
3851 CCACCAGAGT TAACTAATGC CAGTACCCCA TGCCCAACA AGGAATGCCT
3901 TTGGGCTCCA CTGTCAATTT CAGAGCCTCA AAAATAATTC AACCTAGTC
3951 CCTGCTTAAC CCATTAAGCC ACCTAACCAG CAGCTGGGAA ATTCCAGCAT
4001 TGATCTAGA CCCCTGTTAT CCAAGATTGG AGAACAGTGG GACAAAGTGC
4051 TCCTCTCCAC CATTCTGCG TGTCCCTGGG GAAGATGAGC AGAGCAGAGC
4101 CAGACAGTAA AGGAGAGGGC CACGCCCCCT CCACAGTTA CCTCCTTGGT
4151 ACTCCTGCCC GCACTACCCA CAGCAACCCC GGGATGCCGA TCTGCAGCCA
4201 CATGTCCCAT GTGGGAGGTT TCTGCTGAAA GAACTTCCAA CTACACATCT
4251 CCCCACTTCA GTATAAATTT CAACCTTCCC TAATTCATGC AACCTTTTTT
4301 TTTTTTTTTT TTTTTTGAGA CAGAGTGTCG CTCTGTCACC GAGGCTGGAG
4351 TTCAGTGATG CAATCTCGGC TCACTGCAAC CTCTACCTCC TGGGTTCAG
4401 CTATTCCTCT GTCTCCGCTT CCCAAGTAAC TGGGACTACA GCGGTGTGCC
4451 ACCACTCCTG GCTAGTTTTT TGTATTTTTA GTAGAGATGG GGTTCACCT
4501 TGTGGTTCAG GCTGGTCTCA AACTCCCAAC TCAGGTGATC CGTCCACTTG
4551 GGCACCCAAA ATGNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
4601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
4651 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
4701 NNNNNNNNNN NNNNNNNNNN TTCAAGTACC AGCCTGGCCA ACATGGTAGA
4751 AACCCTGCTT CTAATAAATA TAAATAAATA GCCAGGCGAG GTGGTGCATG
4801 CCTATAATCC CAGCTACTCA GGTAGGCTGA GGCAGGAGAA TCATTTAAAC
4851 CTGGGAGGTG GAGGTTGTGG TGAGCCAAGA TCTCGCCATT GCACTCCAGC
4901 CTGGGCAACA AGAGCAAAAC TCCGTCTCAA AAAAAAAG AAAGAAAGAA
4951 AGAAAGAAAC TTCCAAATAA ATGTTGTGAC ACAAAAAA AAACCCAAAC
5001 AATATTCATT ATAGAGTATG CAAATGACCA TGCCCCACCC CCAGCAGATT
5051 CTGATAGACT CCCTTGGGTG GGAATCCTTG TCCAATATAT TGACACTTCC
5101 CTTTCCTGTC AGTATAGCCC AGCCCATGCG TGTACTCACG AGCGGACGAT
5151 GGATGACACA AGTACACAGA GGGACGGAAT CCCTGCATGG TGTGGCTATG
5201 GGCAATGTG GCCACTGTCT AGATTGTGCA AATGTGGTGG TTCTCTGGGG
5251 CCACAGAGCA CACTTGGGGA CCTGTTTCATG GTGAGGTCTC AACTCCGGCC
5301 TCTAGGAACT TGAATGAGGA CAGGAGGGTC AGAGGGAGAG CCTAGGAGGC
5351 TGAGCCAAGG AGCGTGGAGA GGAGAGACAG GGTGAAGGTG GCGGCTGGCT
5401 TTCTGGAAGC AGGTGGCCTT TGGTGCGGTC AGCATTCTGT CCAGCCCCCT
5451 CTTCTCTGAT CCTCTCCATG TGTCTCTCTC CTGGAATCCC AGAAGCTGCC
5501 CCTGACTCCC CATTAACTGC CTCTGCCCT ACCCCCTAGG TGATGCTTCT
5551 GGGAGACACA GCGTTCGCA AAACATGTTT CCTGATCAA TTCAAAGACG
5601 GGGCCTTCCT GTCCGGAACC TTCTAGCCA CCGTCGGCAT AGACTTCAGG
5651 GTGAGGTGGC TGCAGGCACT TGCTTCCAGC AGAGAGCCAG GGCTGTGGCT
5701 CAGGCATGGG GGGTTGCCC CCACCTTGCT CACCCTGGCT CCCAGGGACT
5751 CCCGAGGCTC ATGCCTGGAG GGCACACAAC CCGCTCCCCC AAGACCACAG
5801 AGGTGGCCGG GTCAAAGGAG ACTGGGCAAG GTTGGCTCCT TGCCCAACTA
5851 TAGGATGCAA AAAAATGAGA CTGAGTCTTC GATTCCAGCT CCATTCTGG
5901 GGGACTTCTC CCAAGCAGAG CAGCCGAGG CACGGCATAA GCTGAATATC
5951 TTGGCCCA GAGCCCCGTC TCATTGCTCT CCTACCTGGG CCCCTTTGGA
6001 AAGCCCTCAA AGGTCAATCA GTCTTCTG AGTTCCAGA AAGCACAGCC
6051 CTGCACGGG TTTAAGAGCT GGGCTTGGC CAGGCATGGT GGCTCTTGCC
6101 TGTATTCCTA GCACTTTGGG AGGCCGAAGC GGTCAGATCA CAAGGTCAGG
6151 AGTTTGAGAC CAGCCTGGCC AACATGGTGA AACCCTGTCT CTACTAAAAA
6201 TACAAAAATT AGCCAGGTGT AGTGCACGC TCCTGCAGTC CCAGCTACTC
6251 GGGAGGCTGA GGCAGGAGAA TCGCTCAAT CCGGGTGGTG GAGGTTGCAG

FIGURE 3, page 2 of 7

6301	TGAGCTGAGA	TCGCGCCACT	GCACTCCAGC	CTGGGCAACA	AAGTGAGACT
6351	GCGTCTCAGA	AAAAAAAAAA	AAAAAAGAGC	TGGGCTGGCC	ATGTTGGGAG
6401	ACAGCAGCTC	ACCAGGGACC	CTCCCTCTCA	CCTTGACGAC	TCCATCTTAC
6451	AAATCTGCAT	CAGGGATGCT	AGACGCTGCA	CACCTGAAGT	GTTCAATAGA
6501	GAAGAAGTCT	CACCTTGGCA	GGTGGGGCTC	TACAGCTTCA	AGCAGGCAGA
6551	AAGCGAACAC	TTCCTTCACT	AGAGAATTAG	TGGGCAGCTA	AAGAAAAGGT
6601	GCTGCTGCAG	ATGTAGCCTC	AGGTCCCCAG	GATGCAGGCA	AACACCCCAT
6651	CTCCAGGGGC	TCGTTCACAG	TCCCAAGGCT	AGGCTCCAGG	AGAGGGAGAC
6701	CGAAGTGGGG	AAAGGGCAGG	GCCTCCAGCA	GCAACCAGCC	CTCCAGCCCT
6751	GGGCTGCCTG	ATCCCTGGAG	AGAGCCAGGA	TGTTTCTCAG	GCTCCTCTTG
6801	CCCTGCTGTT	GTGAGAAGGC	AGTTACAGTC	CTCAGAAGGG	ACGACTCCAC
6851	AGTGGAGGTG	TCTGGGTATG	GGGTTCCCTG	TGCCCTGATG	GTATGATCTG
6901	GCTGGAGACG	GTTCTGGGGC	TCACTGCACC	CACTCTAGGC	CTGGAGAGGG
6951	AACAAGAGAG	GACGTCTGCA	GAGCTGAGGA	GCCACATGAC	TCCTGCCCTC
7001	CCATCCTCTG	CCTTTTTCTC	TTTCAGAAC	AGGTGGTGAC	TGTGGATGGC
7051	GTGAGAGTGA	AGCTGCAGGT	GAGACCAGAG	GCTGGAGTTG	GGGAGGGAGG
7101	ATGGAGGACC	TGCCCTTCCT	TCTCACCCTG	AACCACAGGA	GGCCTGCAGC
7151	CCTGCCCTCC	CTCGGGGCA	ATTTCCTGTG	GGGCCACGG	GAGGAAATGG
7201	CTTTTGTTTA	TTTGACATCT	GCAGAAAAAG	CAGTTCCAG	GCACCCCTCTC
7251	ATCTATGAAC	AGCAGCTCCA	AATGCCTTCA	GACAAGCTTA	GCCTCCATCC
7301	ATCTCCTCCC	CAGTTGCCAG	GGCTTTATCT	GCTCTTAGGA	GATTGGACAT
7351	CCCCAACCCC	TGAGCTAGGG	GAGAGGAGAA	GATTCTTTTT	TTTTCTTTTC
7401	TTTTCTTTTT	TTTTTTGAGA	TGGAGTCTCG	CTCTGTCGCC	CAGGCTGGAG
7451	TGCAGTGGCA	CAATCTCGGC	TCACTGCAAC	CTCTGCCTCC	CAGGTTTAAG
7501	AGATTCTCCT	CCCTCAGCCT	CCTGAGTAGC	TGAGACTACA	GGTGCATGCC
7551	ACCACACCTG	GCTAATTTTT	TGTATTTTTA	GTAGAGACGG	GGTTTCACTG
7601	TGTTAGCCAG	GATGGTCTGG	ATCTCCTGAC	CTCGTGATCC	GCCTGCCTCG
7651	GCCTCCCAAA	GTGCTGGGAT	TACAGGTGTA	AGCCACCGCG	CTCGGCTGAG
7701	GAGATGATTT	TGAACGAGCT	TGAGAAATCA	GTAAC TGCTA	CTGTCCAGGT
7751	CATTGGATGC	TCAGGGGCTC	ATGAGAACCT	AAAGAAGAAA	ACAGCCCCAC
7801	TTTCCCACAG	ATATCTCATA	CAACAAAGCA	GGCCTGCTCC	ACCCAGCACA
7851	CTTCTTGAC	GTGCTCCTT	CTGACCATTT	CTCCATCCCA	TCCCTTCCCA
7901	GATCTGGGAC	ACCGCTGGGC	AGGAACGGTT	CCGAAGCGTC	ACCCATGCTT
7951	ATTACAGAGA	TGCTCAGGGT	GAGTCCCTCG	CACCTCCAA	CCCCTACCCC
8001	AGCCCCTTGG	TAGCATCCGT	GCTGCTGCCT	AAGTCCCCTC	TGTGATCCTC
8051	TCCCCTCCAG	CCTTGCTTCT	GCTGTATGAC	ATCACCACAA	AATCTTCTTT
8101	CGACAACATC	AGGGTAGGTC	CTCCCTTCCC	CTGACTCCCA	CCCATAAGCA
8151	GCCAAGGCAA	GGTCTATGCA	GGCTGGGGTT	GCTTCCTGCC	CTGTGGAAAG
8201	CGGGTGGAGC	GTGGAGTCC	CCTGCCTTCT	GAAAAACACC	TACTTGTGAC
8251	TCAGAAGTCA	TATCTGCTGC	TTTGTATTTG	GTGGCCATGT	GGGCATGAAG
8301	GCCAAGCAGG	CTGTTGTGAC	CCTGTGCCAC	CTGCATAGCC	CTCACTGTGA
8351	TTACAGAGTG	TGTTTCGTGA	CAAAGTGTTC	AGAACAGCCC	CCACTCCACC
8401	CTGGATAATT	ATCCACAGAG	ACCAAGGGAA	AAACACAACC	AGAAAAGTCC
8451	ACACATACAT	CCAGGGCAAG	TTGCAAGAAA	GTGACTCAGT	CAGACAGAGT
8501	GAGTGGTTGT	ATCTCACA	CCAACTATT	ATAGAGACAA	AAATTTGATA
8551	AATTCAAGCA	CCAATTTTGT	TCACGACATT	GTATAGGTTT	CATGAATCCC
8601	CTGACCTCAA	GGACAGTTTG	CTGATAAGCA	AACTAGGAGA	ATAAAACGTT
8651	TATATAGAAA	GAGGAAAATC	CATGGCACTC	ATACTCCTAC	CTCCAACCCC
8701	ATGCTCATGG	CAGACATCAC	TAATCAATCA	CAGTACTTTT	GATCACTGAA
8751	ACCCTTATGT	GGTCTTAGAA	TCTTTAACAG	GACACTCCAA	GAAATCACTG
8801	CTGACAGCCA	ACTGATTTGT	GAGATAAGGT	CTCCATGCAT	CTGGATCTTC
8851	CATAGAAGTG	ATAGTTGCAC	AGCATAAAAT	GGTGAGGGTG	GGGCCATTGT
8901	GGGTTGAGCC	ACCAAGGAAG	GCCATCCAGG	CCTGGATGGG	CCAGAACAAA
8951	GGTACAGATG	AGAGAACGCA	CAGGGTATCG	TGTTCAAGGT	AGTGAGTAAC
9001	TGAGGATAGT	CAAACGGAGC	AGAAGAAGAA	AGGGGCAGCA	GGAGGAAGAG
9051	AATGCCAGTC	TCGCACGCCC	TCTCCCACAG	GCCTGGCTCA	CTGAGATTCA
9101	TGAGTATGCC	CAGAGGGACG	TGGTGATCAT	GCTGCTAGGC	AACAAGGTGA
9151	GTGGCTCCGG	GGCAGGGTCA	GCCCAGCCCT	GCACTTCCTC	AGCCCTAGCC
9201	GGCCCCATAA	CCACCCAAGA	ACAGTTATCT	AGGCATCCTT	CCTGAAAAGG
9251	ACTCTGCAGC	CTCCAGCTCA	GGGGTCAGAC	ATATCTGGAG	GCTTCTGCC
9301	ATCCCATCTG	CCCCTTCCAG	GGAAAGTCCA	AGTTGTTGCC	TGAGAAATCA
9351	AGGGGTGCCC	AGTTCTCAGC	CCCCATTAGA	GCAGAGTGAA	CAGGGTCCCA
9401	GGTCAGGGGC	TAAGAGTGCA	AAGGGTTAGC	CCCAACTGCT	GTCCTATTCC

FIGURE 3, page 3 of 7

9451 AAGACCCTTT ACCAAAGGTG AGATCCCAGA GCTGGGAGCT ACACTGGGCA
9501 GAAACCCTGG CCCAGGCCA ATCACACCTG CCTGCAGTCC CTTGGGCCAC
9551 CAGCAGAGGG CAGGCAACGC CTGCTTCTGG GGCAAAATAT GGGCCCGCTG
9601 GGGCGGAGGC CTCCTTCCCC AGAGTGACCC ATTTGGGCTT GACAGGCGGA
9651 TATGAGCAGC GAAAGAGTGA TCCGTTCCTG AGACGGAGAG ACCTTGGCCA
9701 GGGTAAGTGA TTGTCTGTGG GACAGGGTGA AGGGTGGGGG CAACCCGACG
9751 CTGGCCCTGA GGACACTCTC TCCCGGGCAG GAGTACGGTG TTCCCTTCCT
9801 GGAGACCAGC GCCAAGACTG GCATGAATGT GGAGTTAGCC TTTCTGGCCA
9851 TCGCCAAGTG AGAGCTGGGC AGGGAAGGGA AGTGTGCGGG GCAGGGCGGC
9901 ACACTCCAGG AATCCAGTAG GGCCCGGCC CTGGCCAGC CCCTGGACAC
9951 ACCTGCATTC TGCAAGCTGA GGTCCATTG CTCTGGGAGC ACTGGGCCAC
10001 TGGGAGAGGG GAGGGGGCGG CTCAGCTCCT CACCCAGCC CAGCCAGCC
10051 CAGCCAGCC CATTGTCTCT TCTTCAAGGG AACTGAAATA CCGGGCCGGG
10101 CATCAGGCGG ATGAGCCAG CTTCAGATC CGAGACTATG TAGAGTCCCA
10151 GAAGAAGCGC TCCAGCTGCT GCTCCTTCAT GTGAATCCCA GGGGGCAGAG
10201 AGGAGGCTCT GGAGGCACAC AGGATGCAGC CTTCCCCCTC CCAGGCCTGG
10251 CTTATTCCAA GAGGCTGAGC CAATGGGGAG AAAGATGGAG GACTCACTGC
10301 ACAGCCGCTT CTTAGCAGGG AGCTATACTC CAACTCCTAC TTGAGTTCTT
10351 GCGGTCTCCC CGCATCCACA GGGAGGGTAA AACACTTAGC TTTTATTTTA
10401 ATAGTACATA ATTTAATACC AAAAAAGGCG CCTGGATCCC CAAAAACCG
10451 AGGTGGGAG CTAGTGGCCC TTTTGCTTTC TAGGACTTGG GGGGCCGGCC
10501 CTCCTCCTA AGCATAACAA AGGTGGTGTG GCTCCAGCTC AGCCCCAGGG
10551 GACACAGATG CACTTTGGGG GTGAGGGCAG GTAATGACTC CATCGCACCC
10601 TCAGTTCAGC TGGACAGAGG CTCAGGTGAC CCCAGCCTTC ACTGTCTCCC
10651 GCTCTCCAGG AGCTTATCTT CGCCCCATCT CCCAAATAAG TGGGCCCTTG
10701 TGCTGTGAGG AAGACCAAAG CCTCAGGGAA GATAAGAGAT ATGGAGATGG
10751 GAGGGGGAGG ACAAGGGGCA GAGAGTAGGG TCTAGCTGGC TATCTCTGGC
10801 CTTACTAACA CCCCCCTGGA GGCATGCCCC TTTTCTCCAG CACACAAGCA
10851 CATTGGGGCA CCTGGAAATA TTGGTTCCAG GCTCCTGTTC TCTGGACTTC
10901 AGATCCTGGG GGAGCCCTC CCCCCCTGA ATCCCTGGCT TAGCTACCTT
10951 CCTGCCTGTG CACCTAAAAA CCTCAGGTCA GAACTAGGAA AAGAGTTTTG
11001 TTTTATTTT TTTGAAATGG AGTCTCGTTC TGTGCCCCAG GCTGAGGTGC
11051 AGTAGTCAA TCTCCGCTCA CTACAACCTC CACTCCCTGG GGCTCAAGCG
11101 ATCTCCAC ATCAGCCGCC GAAGTAGCTG GGACTATAGG TGTGTACCAT
11151 CACACCTGGC TAATTTTTGT ATTTTTTGTA GACACAGGGT TTCGCCATGT
11201 TGCCAGGCT GGTCTTGAAT TCCTGAGCTC AAGCAACCTG CCGGCCTCGG
11251 CCTCCCAAAG TACTGGGATT ACACGCAGAA GGCACCATGC CCAGGCTAGA
11301 TGTGTCTTAT CCCAATCCTT TGGCAGGCAT GCAGCTCCAC AGGCGATTTT
11351 TTCAAGCAGC TGAAGTGTTC AGCCCTCCTG GGTAAAGAGC CAGATAAGGA
11401 GAAATCCCTT TCCTAGGTTT GTGAAAAAAA AGAGAAATCC
11451 CTGGCTCCTG GAGCTGGTGG GAGACAAGAT TAAGCAAACC TCCCTGACA
11501 TGTATCCCTT TGACCCCAAG CTCTGCCTCC TCCCTGACCA CCCATGCCCT
11551 TTCTTTTAA TTTCTCAAACA GATACCAGGG CCTAAACTGC TTTACCTCCC
11601 CTCCTACTGA GTCAGGTTAG GTGGTGGGAG GTCACCCATT TCCGAGTTAA
11651 ACCAATGCAA TATGAGTAAA ACAAAGTCAT GTGGGTATGT CTGGGGTAGA
11701 GAGAGGGGTA TCAAGTTTCT GTGTCTCCTT TGATCACATA TCTCCCAAAG
11751 CTCTGATCCC TGCCATGGGA AGTGGACAGG AAACATGAGG TCATGACCTG
11801 CAGGCATCTT TACTGCAGCT CTGCCGCTC GGAGGGGGAG AGGGGGAGGA
11851 AGAAGTATGC GCTGCACATT TCTGAGGCTA CTGCATTTGC TTTCAAGGCA
11901 GAAATCTTGC TCTGAGCAGT CAGCGCTCC AGTTTGGGCC CGATAAGGAA
11951 GTTCTCCGTG GCCTCCCTCA GGCAGAGCAG GGAGGAGGCT GACATTGCCA
12001 GTCTCTTCTG GCGCCCAAGG CAGGTTGCAG GAGATCCAAT CCCATAGACA
12051 GCTCTGGGCC TCTTGCAATT GAGTTTTTCA GAATTAAACT GCAGTATTTT
12101 GGAAAGCACA TCCTGTCCAC TGTCTCTTTG AAGTGAGTGG GGGGGGGGGG
12151 TCTTGTTGAA GGAATTGTCA TTCCTGCTCA AAATCATTCC ATCCTCCTTC
12201 CTCAGTGTCT GTCCTCAGAT GGTCAGCTCC CCGCTCAACA GACTGTCTCC
12251 CGCCTCTGTG ACCAGCCTCT CTTTGCCAAG AGGGAGCTAG AAGGCTTTAC
12301 AGTCCTAATC ATTTTCTGTG TGGAAAAAAA AAAAAAAAC CAAGGCTCCT
12351 TTCCCTGTGG CGTGTACCA GAGGTGATT ACCTGAGTCT GTCCTGCCTC
12401 TCCCCACCCC ACCTCCCTAG CCAAACGCTG CTGCCAAAGC CCACGCTATT
12451 GCCCTAGATG GCCTGTCTTC AGCGGGCTGC CCCTCGAGGT CCCAGGCTCT
12501 CCGCGGAGCC CTCACCTTCC CAGCAGGGAT CAGAACCTGC ACTCCTCTAT
12551 GCGAGTCCTG GGACAGCACA AAGTGATTA GGGTTAGGGT TCCACAAAC

FIGURE 3, page 4 of 7

12601 GGAAAAATGT TATTCAAACA ACTCTGTAGG GTCCGAGGAG GCCCTCCGTC
12651 TTAATTCTCG AGACTGACCG GCCCTCGCTG CCCCAGCGCG GAGCAGTTGC
12701 CCCGGCAACA GCCGCTCCCT CTCAACTGGA GCTGCACCCA GGCTTTGGCT
12751 AAAGGCTGTT AAAACGTTGG CCAGGTGCGG AGGCTCACGT CTGTAATCCC
12801 AGGGCGGATC ACCTGAGGTC AGGAGTTTGA AACCATCCTG GCCAACATGG
12851 CGAAATTTTCG TCTCTACTAA AAATACAAAA ATTAGCGGGG CGTGGTGGTG
12901 CGCGCCTGTA ACCCCAGCTG CTCGGGAGGC TGAGGCAGGG GAATCGCTTG
12951 AACC CGGGAG GCGGAGGTTG CAGTGATCCG AGATCGCGCC ACGGCAGTCC
13001 AGCCTGGGCG ACAGAGCGAG ACTCCGTCTC AAAAAAAAAA AAAAAAGTTA
13051 GGGTCCTTTA CCCGAGGGCC GGCTTTCCTC ACTCCCCGCC ACAGGTAGGG
13101 GAAACCAGGC CGGAGCCGCG GGGCCCACCC GCCCAGAACC GGAATTTCGG
13151 CGAGCCCCGC CCCTGCCACC CCAGCGCCGG CC (SEQ ID NO:3)

FEATURES:

Start: 2042
Exon: 2042-2134
Intron: 2135-5539
Exon: 5540-5650
Intron: 5651-7026
Exon: 7027-7068
Intron: 7069-7901
Exon: 7902-7968
Intron: 7969-8060
Exon: 8061-8113
Intron: 8114-9080
Exon: 9081-9146
Intron: 9147-9645
Exon: 9646-9702
Intron: 9703-9780
Exon: 9781-9857
Intron: 9858-10078
Exon: 10079-10181
Stop: 10182

CHROMOSOME MAP POSITION:

Chromosome # 17

ALLELIC VARIANTS (SNPs):

DNA			
Position	Major	Minor	Domain
4259	C	T	Intron
4325	G	T	Intron
4348	G	A	Intron
4924	G	A	Intron
4983	-	A	Intron
6710	A	G	Intron
8624	A	G	Intron
8661	G	A	Intron
11754	T	C	Beyond ORF(3')
11836	A	G	Beyond ORF(3')

Context:

DNA

Position

4259

ACCCATTAAGCCACCTAACCCAGCAGCTGGGAAATTCAGCATTGGATCTAGACCCCTGTT
ATCCAAGATTGGAGAACAGTGGGACAAAGTGCTCCTCTCCACCATTCCTGCGTGTCCCTG
GGGAAGATGAGCAGAGCAGAGCCAGACAGTAAAGGAGAGGGCCACGCCCCCTCCACAGGT
TACCTCCTTGGTACTCCTGCCCCGACTACCCACAGCAACCCCGGGATGCCGATCTGCAGC
CACATGTCCCATGTGGGAGGTTTCTGCTGAAAGAACTTCCAACCTACACATCTCCCCACTT
[C, T]
AGTATAAATTTCAACCTTCCCTAATTCATGCAACCTTTTTTTTTTTTTTTTTTTTTTTGAG
ACAGAGTGTGCTCTGTACCGAGGCTGGAGTTCAGTGATGCAATCTCGGCTCACTGCAA
CCTCTACCTCCTGGGTTCAAGCTATTCTCCTGTCTCCGCTCCCAAGTAAGTGGGACTAC
AGGCGTGTGCCACCACTCCTGGCTAGTTTTTTGTATTTTGTAGTAGAGATGGGGTTTCACC
TTGTTGGTCAGGCTGGTCTCAAACCTCCCAACTCAGGTGATCCGTCCACTTGGGCACCCAA

4325

GATTGGAGAACAGTGGGACAAAGTGCTCCTCTCCACCATTCCTGCGTGTCCCTGGGGAAG
ATGAGCAGAGCAGAGCCAGACAGTAAAGGAGAGGGCCACGCCCCCTCCACAGGTTACCTC
CTTGGTACTCCTGCCCCGACTACCCACAGCAACCCCGGGATGCCGATCTGCAGCCACATG
TCCCATGTGGGAGGTTTCTGCTGAAAGAACTTCCAACCTACACATCTCCCCACTTCAGTAT
AAATTTCAACCTTCCCTAATTCATGCAACCTTTTTTTTTTTTTTTTTTTTTTTGAGACAGA
[G, T]
TGTCGCTCTGTACCGAGGCTGGAGTTCAGTGATGCAATCTCGGCTCACTGCAACCTCTA
CCTCCTGGGTTCAAGCTATTCTCCTGTCTCCGCTCCCAAGTAAGTGGGACTACAGGCGT
GTGCCACCACTCCTGGCTAGTTTTTTGTATTTTGTAGTAGAGATGGGGTTTCACCTTGTG
GTCAGGCTGGTCTCAAACCTCCCAACTCAGGTGATCCGTCCACTTGGGCACCCAAAATG

4348

TGCTCCTCTCCACCATTCCTGCGTGTCCCTGGGGAAGATGAGCAGAGCAGAGCCAGACAG
TAAAGGAGAGGGCCACGCCCCCTCCACAGGTTACCTCCTTGGTACTCCTGCCCCGACTAC
CCACAGCAACCCCGGGATGCCGATCTGCAGCCACATGTCCCATGTGGGAGGTTTCTGCTG
AAAGAACTTCCAACCTACACATCTCCCCACTTCAGTATAAATTTCAACCTTCCCTAATTC
TGCAACCTTTTTTTTTTTTTTTTTTTTTTTGAGACAGAGTGTGCTCTGTACCGAGGCTG
[G, A]
AGTTCAGTGATGCAATCTCGGCTCACTGCAACCTCTACCTCCTGGGTTCAAGCTATTCTC
CTGTCTCCGCTCCCAAGTAAGTGGGACTACAGGCGTGTGCCACCACTCCTGGCTAGTTT
TTTGTATTTTGTAGTAGAGATGGGGTTTCACCTTGTGGTCCAGGCTGGTCTCAAACCTCCA
ACTCAGTGATCCGTCCACTTGGGCACCCAAAATG

4924

TTCAAGTACCAGCCTGGCCAACATGGTAGAAACCCCGTCTCTACTAAAAATAAAAAATTA
GCCAGGCGAGGTGGTGCATGCCTATAATCCCAGCTACTCAGGTAGGCTGAGGCAGGAGAA
TCATTTAAACCTGGGAGGTGGAGGTTGTGGTGAAGCAAGATCTCGCCATTGCACTCCAGC
CTGGGCAACAAGAGCAAAACTCC
[G, A]
TCTCAAAAAAAAAAGAAAGAAAGAAAGAAAGAACTTCCAAATAAATGTTGTGACACAA
AAAAAAAAACCCAAACAATATTCAATTATAGAGTATGCAAATGACCATGCCCCACCCCA
CAGATTTCTGATAGACTCCCTTGGGTGGGAATCCTTGTCCAATATATTGACACTTCCCTTT
CCTGTCAGTATAGCCAGCCCATGCGTGTACTCACGAGCGGACGATGGATGACACAAGTA
CACAGAGGGACGGAATCCCTGCATGGTGTGGCTATGGGCAAAATGTGGCCACTGTCTAGAT

4983

TTCAAGTACCAGCCTGGCCAACATGGTAGAAACCCCGTCTCTACTAAAAATAAAAAATTA
GCCAGGCGAGGTGGTGCATGCCTATAATCCCAGCTACTCAGGTAGGCTGAGGCAGGAGAA
TCATTTAAACCTGGGAGGTGGAGGTTGTGGTGAAGCAAGATCTCGCCATTGCACTCCAGC
CTGGGCAACAAGAGCAAAACTCCGTCTCAAAAAAAAAAGAAAGAAAGAAAGAAAGAAAC
TTCCAAATAAATGTTGTGACAC
[-, A]
AAAAAAAAACCCAAACAATATTCAATTATAGAGTATGCAAATGACCATGCCCCACCCCA
GCAGATTCTGATAGACTCCCTTGGGTGGGAATCCTTGTCCAATATATTGACACTTCCCTT
TCCTGTCAGTATAGCCAGCCCATGCGTGTACTCACGAGCGGACGATGGATGACACAAGT
ACACAGAGGGACGGAATCCCTGCATGGTGTGGCTATGGGCAAAATGTGGCCACTGTCTAGA
TTGTGCAAAATGTGGTGGTTCTCTGGGGCCACAGAGCACTTGGGGACCTGTTTCATGGT

6710

CACCAGGACCCCTCCCTCTCACCTTGACGACTCCATCTTACAAATCTGCATCAGGGATGC
TAGACGCTGCACACCTGAAGTGTTCAATAGAGAAAAGGTCTCACCCTGGCAGGTGGGGCT

FIGURE 3, page 6 of 7

CTACAGCTTCAAGCAGGCAGAAAGCGAACACTTCCTTCACTAGAGAATTAGTGGGCAGCT
AAAGAAAAGGTGCTGCTGCAGATGTAGCCTCAGGTCCCAGGATGCAGGCAAACACCCCA
TCTCCAGGGGCTCGGTACAGTCCCAAGGCTAGGCTCCAGGAGAGGGAGACCGAAGTGGG
[A, G]
AAAGGGCAGGGCCTCCAGCAGCAACCAGCCCTCCAGCCCTGGGCTGCCTGATCCCTGGAG
AGAGCCAGGATGTTTCTCAGGCTCCTCTTGCCCTGCTGTTGTGAGAAGGCAGTTACAGTC
CTCAGAAGGGACGACTCCACAGTGGAGGTGTCTGGGTATGGGGTTCTGCTGCCCTGATG
GTATGATCTGGCTGGAGACGGTTCTGGGGCTCACTGCACCCACTCTAGGCCTGGAGAGGG
AACAAGAGAGGACGTCTGCAGAGCTGAGGAGCCACATGACTCCTGCCCTCCCATCCTCTG

8624 GTGCCACCTGCATAGCCCTCACTGTGATTACAGAGTGTGTTTCGTGACAAAGTGTTTCTGAC
ACAGCCCCCACTCCACCTGGATAATTATCCACAGAGACCAAGGGAAAAACACAACCAGA
AAAGTCCACACATACATCCAGGGCAAGTTGCAAGAAAGTGACTCAGTCAGACAGAGTGAG
TGGTGTGATCCTCACAACCAAACTATTATAGAGACAAAAATTTGATAAATTCAAGCACCA
ATTTTGTTCACGACATTGTATAGGTTTCATGAATCCCCTGACCTCAAGGACAGTTTGCTG
[A, G]
TAAGCAAACCTAGGAGAATAAAACGTTTATATAGAAAGAGGAAAATCCATGGCACTCATAC
TCCTACCTCCAACCCCATGCTCATGGCAGACATCACTAATCAATCACAGTACTTTTGATC
ACTGAAACCTTATGTGGTCTTAGAATCTTTAACAGGACACTCCAAGAAATCACTGTCTGA
CAGCCAACCTGATTTGTGAGATAAGGTCTCCATGCATCTGGATCTTCCATAGAAGTATAG
TTGCACAGCATAAAATGGTGAGGGTGGGGCCATTGTGGGTTGAGCCACCAAGGAAGGCCA

8661 TGTTTCGTGACAAAGTGTTTCTGACAAAGTGTTTCTGACAAAGTGTTTCTGACAAAGTGTTTCTGAC
ACCAAGGGAAAAACACAACCAGAAAAGTCCACACATACATCCAGGGCAAGTTGCAAGAAA
GTGACTCAGTCAGACAGAGTGAGTGGTGTGATCCTCACAACCAAACTATTATAGAGACAA
AAATTTGATAAATTCAAGCACCAATTTTGTTCACGACATTGTATAGGTTTCATGAATCCC
CTGACCTCAAGGACAGTTTGCTGATAAGCAAACTAGGAGAATAAAACGTTTATATAGAAA
[G, A]
AGGAAAATCCATGGCACTCATACTCCTACCTCCAACCCCATGCTCATGGCAGACATCACT
AATCAATCACAGTACTTTTGATCACTGAAACCTTATGTGGTCTTAGAATCTTTAACAGG
ACACTCCAAGAAATCACTGCTGACAGCCAACCTGATTTGTGAGATAAGGTCTCCATGCATC
TGGATCTTCCATAGAAGTATAGTTGCACAGCATAAAATGGTGAGGGTGGGGCCATTGTG
GGTTGAGCCACCAAGGAAGGCCATCCAGGCCTGGATGGGCCAGAACAAAGGTACAGATGA

11754 GCTCCTGGAGCTGGTGGGAGACAAGATTAAGCAAACCTCCCCTGACATGTATCCCTTTGA
CCCCAAGCTCTGCCTCCTCCCTGACCACCCATGCCCTTTTCTTAACTTCTCAAACAGAT
ACCAGGGCCTAAACTGCTTTACCTCCCCTCCTACTGAGTCAGGTAGGTGGTGGGAGGTC
ACCCATTTCCGAGTTAAACCAATGCAATATGAGTAAACAAAGTCATGTGGGTATGTCTG
GGGTAGAGAGAGGGGTAGCAAGTTATGTGTCTCCTTGGTACATATCTCCCAAAGCTC
[T, C]
GATCCCTGCCATGGGAAGTGAGCAGGAAACATGAGGTGATGACCTGCAGGCATCTTTACT
GCAGCTCTGCCGGCCTGGAGGGGGAGAGGGGGAGGAAGAAGTATGCGCTGCACATTTCTG
AGGCTACTGCATTTGCTTTCAAGGCAGAAATCTTGCTCTGAGCAGTCAGCGGCTCCAGTT
TGGGCCCCGATAAGGAAGTTCTCCGTGGCCTCCCTCAGGCAGAGCAGGGAGGAGGCTGACA
TTGCCAGTCTCTTCTGGGGCCCAAGGCAGGTTGCAGGAGATCCAATCCCATAGACAGCTC

11836 GACCACCCATGCCCTTTCTTAACTTCTCAAACAGATACCAGGGCCTAAACTGCTTTAC
CTCCCCCTCTACTGAGTCAGGTTAGGTGGTGGGAGGTCACCCATTTCCGAGTTAAACCAA
TGCAATATGAGTAAACAAAGTCATGTGGGTATGTCTGGGGTAGAGAGAGGGGTAGCAAG
TTCATGTGTCTCCTTGGTCACATATCTCCCAAAGCTCTGATCCCTGCCATGGGAAGTGG
ACAGGAAACATGAGGTCATGACCTGCAGGCATCTTTACTGCAGCTCTGCCGGCCTGGAGG
[A, G]
GGAGAGGGGGAGGAAGAAGTATGCGCTGCACATTTCTGAGGCTACTGCATTGCTTTCAA
GGCAGAAATCTTGCTCTGAGCAGTCAGCGGCTCCAGTTTGGGCCCCGATAAGGAAGTTCTC
CGTGGCCTCCCTCAGGCAGAGCAGGGAGGAGGCTGACATTGCCAGTCTCTTCTGGGGCCC
AAGGCAGGTTGCAGGAGATCCAATCCCATAGACAGCTCTGGGCCTCTTGCATTTGAGTTT
TTCAGAAATTAAGTGCAGTATTTTGGAAAGCACATCCTGTCCACTGTTTCTTTGAAGTGA

FIGURE 3, page 7 of 7